

The Relationship Between Student Performance and Instructor Evaluations Revisited

R. Eric Landrum and Ronna J. Dillinger
BOISE STATE UNIVERSITY

ABSTRACT

Students in introductory psychology completed an end-of-semester evaluation containing specific and global questions concerning instructor performance and course evaluation. Students' actual and expected course grades were matched with evaluation outcomes. Global items referring to overall course and instructor were significantly correlated. Whereas the instructor evaluation is weakly (but significantly) correlated with actual grade (but not with expected grade), the course evaluation is not significantly correlated with actual grade (but is weakly yet significantly correlated with expected grade). The results are discussed in the context of the differential predictors for course and instructor evaluation.

& Ogasawara, 1981), type of class (DaRosa, Kolm, Follmer, Pemberton, Pearce, & Leapman, 1991; Hoffman, 1983) and difficulty level of the class (Marsh, 1978; Schwab, 1975) can all affect course ratings (Centra, 1993; Ellis & Rickard, 1977; Marsh, 1978). Teacher characteristics, such as attire (Chowhardy, 1988), how animated the instructor speaks (Centra, 1993; Williams & Ceci, 1997), personality (Centra, 1993; Kovacs & Kapel, 1976), and instructor sexual orientation (Liddle, 1997) can also potentially influence ratings. For example, Williams and Ceci (1997) conducted an experiment to examine how a specific teaching style affected ratings. During the first semester, the instructor taught in a normal manner. During the second semester, however, the instructor consciously exhibited more enthusiasm during lectures by using wider pitch variation in voice and more gestures. Student evaluation scores were reliably higher in the second semester. It is important to note that student performance in both classes was not significantly different. The conclusion of Williams and Ceci (1997) is that while student learning remained unchanged, students did give higher evaluations based on teaching style. If an instructor can significantly raise ratings simply by becoming more animated without increasing student learning, are the ratings meaningful?

The relationship between actual grades and instructor evaluations has also been examined. Again, the results are conflicting. Some research concludes that the relationship between actual grades and evaluations is reciprocal (Gigliotti & Buchtel, 1990; Hoffman, 1983; Kennedy, 1975; Weigel et al., 1971; Worthington & Wong, 1979). The study by Worthington and Wong, for example, manipulated grades in order to examine the effects of grades. Groups of students with manipulated grades were compared to students who did not receive manipulated grades. Those who received higher grades gave higher evaluations. Marsh and Roche (1997) have criticized this study for the use of deception, as well as circumstances when the manipulated assigned grade was vastly different than the one the student was likely to expect.

The current study examined the ability of students to make a distinction between the evaluation of the course and the evaluation of the instructor. Is the grade a student receives related to that student's evaluation of the course and/or the instructor?

INTRODUCTION

Using instructor evaluation instruments in the college setting is a widespread practice in the United States (Centra, 1993; Greenwald & Gilmore, 1997; Ory & Parker, 1989; Wilson, 1998). Evaluations are used not only for instructor or course improvement, but often as an essential component of administrative decisions (Centra, 1993; Moomaw, 1977). Past research has examined possible influences on evaluations, as well as whether evaluations are an effective method for students to share their impressions of the course and of the teaching with the instructor (Greenwald & Gilmore, 1997; Hoffman, 1983; Kovacs & Kapel, 1976; Marsh & Roche, 1997). It has also been suggested that the deliberate inflation of grades or liberal grading by instructors can be an indirect contributor to higher evaluations (Hoffman, 1983; Vasta & Sarmiento, 1979; Weigel, Oetting, & Tasto, 1971; Worthington & Wong, 1979). This raises certain ethical issues with regards to teaching practices – that is, instructors may be tempted to give higher grades for personal achievement through promotion or tenure via higher evaluations (Jensen, 1987).

Many factors can be potential confounds of instructor ratings. Course characteristics, such as class size (Meredith

METHOD

Participants

The participants in this study were students enrolled in two sections of the first author's General Psychology course ($N = 333$). Participation in the course evaluation process is voluntary. As part of a larger, college-wide evaluation project, students completed two evaluation forms about the first author's course and instruction, and included the last four digits of their Social Security Number (SSN) on both forms so that the evaluation forms could be examined. The results of the newer evaluation form constitute the evaluation analyzed in the present study. This identifier allowed the authors the ability to examine the relationship between student performance in the class, and the student's evaluation of the course and instructor. This process took place well after the semester was complete; in no way were student evaluations examined prior to the assignment of a final grade. Students were advised of this procedure.

Materials

The evaluation questions given to students to complete are presented in Table 1. Note that the first ten items are statements to which participants respond with a Likert-type agreement scale, with 0 = strongly disagree, 1 = disagree, 2 = uncertain, 3 = agree, and 4 = strongly agree. Participants were instructed to leave a question blank if they did not understand the question. Four other questions comprise the evaluation scale. Students were asked about their expected grade (A, B, C, D, or F), an overall rating of the course (excellent, good, fair, poor), the work they performed in the class compared to their classmates (distinguished, superior, average, below average, failure), an overall instructor rating (excellent, good, fair, poor), and the actual grade of students (A, B, C, D, or F).

TABLE 1

Descriptive Outcomes for Course and Instructor Evaluation

<i>Question</i>	<i>Mean</i>	<i>Standard Deviation</i>
1. The instructor's presentation increased my knowledge of the subject.	3.58	0.59
2. The instructor's methods of evaluation were fair.	3.35	0.87
3. The instructor was available during office hours.	3.06	0.90
4. I would recommend this instructor to another student.	3.63	0.74
5. I felt free to participate (e.g., ask questions) in this class.	3.17	0.91
6. The instructor seemed well prepared for class.	3.83	0.42
7. The instructor expressed ideas clearly.	3.60	0.65
8. The objectives of the course were met.	3.56	0.66
9. The assignments and exams were returned in a timely fashion.	3.55	0.80
10. The assignments were of value to my learning.	3.24	0.83
11. I expect to receive a grade of (A = 4, B = 3, C = 2, D = 1, F = 0)	2.77	0.80
12. Overall, I would rate this course as (Excellent = 3, Good = 2, Fair = 1, Poor = 0)	2.55	0.69
13. Compared to that of my classmates, the work I performed in this class was (Distinguished = 4, Superior = 3, Average = 2, Below Average = 1, Failure = 0)	2.38	0.60
14. Overall, I would rate this instructor as (Excellent = 3, Good = 2, Fair = 1, Poor = 0)	2.73	0.57
15. Actual Letter Grade (A = 4, B = 3, C = 2, D = 1, F = 0)	2.21	1.16

Notes: $N = 333$. Items 1-10 are statements to which participants respond with:
0 = strongly disagree, 1 = disagree, 2 = uncertain, 3 = agree, and 4 = strongly agree.

Procedure

About two weeks prior to the end of the General Psychology course, students were given the opportunity to complete the course and instructor evaluation. This process was completed during regularly scheduled lecture time. Students were instructed about the college-wide project to change teaching evaluations and were asked to complete both forms and to put the last 4 digits of their SSN on both forms so that the forms could be looked at together after the semester was complete and after grades were submitted. In addition,

students were informed that after the semester and grades were submitted, their instructor would look at the relationships between their performance in the class and the questions on the evaluation form. Completed evaluations were collected by teaching assistants and delivered directly, in a sealed envelope, to the department secretary. While students were given the entire class period to complete the evaluation forms, most students finished within 30 minutes.

TABLE 2
Intercorrelation Matrix of Instructor Evaluation Items

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. The instructor's presentations increased my knowledge of the subject.	1.000														
2. The instructor's methods of evaluation were fair.	.417**	1.000													
3. The instructor was available during office hours.	.369**	.203**	1.000												
4. I would recommend this instructor to another student.	.396**	.405**	.140	1.000											
5. I felt free to participate in this class.	.328**	.205**	.281**	.242**	1.000										
6. The instructor seemed well prepared for class.	.390**	.179*	.205**	.230**	.218**	1.000									
7. The instructor expressed ideas clearly.	.590**	.298**	.102	.452**	.327**	.429**	1.000								
8. The objectives of the course were met.	.537**	.362**	.371**	.472**	.358**	.454**	.559**	1.000							
9. Assignments & exam results returned in timely fashion.	.347*	.245**	.296*	.273**	.388**	.374**	.324**	.411**	1.000						
10. The assignments were of value to my learning.	.561**	.329**	.228**	.307**	.243**	.272**	.452**	.454**	.402**	1.000					
11. I expect to receive the grade of	.167**	.028	.048	.095	.083	-.032	.024	.094	.078	.140*	1.000				
12. Overall, I would rate this course as	.487**	.375**	.223**	.591**	.188**	.263**	.491**	.453**	.300**	.381**	.207**	1.000			
13. Compared to that of my classmates, the work I performed in class was	.163**	.055	.018	.115	.160*	.102	.109	.052	.016	.110	.414**	.063	1.000		
14. Overall, I would rate this instructor as	.419**	.294**	.183*	.505**	.178**	.285**	.400**	.363**	.314**	.243**	.117	.632**	.048*	1.000	
15. Numerical Letter Grade	.016	-.054	.045	.189*	.067	.022	.079	.135	.013	.037	-.051	.082	.003	.186*	1.000

Notes: Items 1-10 are statements to which participants respond: 0 = strongly disagree, 1 = disagree, 2 = uncertain, 3 = agree, and 4 = strongly agree. Items 11 and 15: A = 4, B = 3, C = 2, D = 1, F = 0.

Items 12 and 14: Excellent = 3, Good = 2, Fair = 1, Poor = 0. Item 13: Distinguished = 4, Superior = 3, Average = 2, Below average = 1, Failure = 0. * indicates $p < .05$. ** indicates $p < .01$.

RESULTS

Descriptive Outcomes

The questions used in this study are presented in Table 1, including means and standard deviations. Results from all of the evaluative questions are presented for completeness. Inter-item reliability analyses were conducted on the first ten items of the evaluation that are similarly scaled. Analyses indicated a Cronbach's $\alpha = 0.85$, indicating very good reliability. Validity analyses were initiated using a factor analysis approach. The same ten items were subjected to a factor analysis using a varimax rotation, eigenvalue cutoff > 1.0 , and factor loadings $> .50$. All ten items loaded on a single factor, with an eigenvalue = 4.63 explaining 46.34% of the variance. The first ten items were selected for the factor analysis procedure because they share the same metric.

Correlational Relationships and Predictors of Instructor Ratings

Correlation coefficients were calculated between all of the items on the instructor evaluation (see Table 2). For the 10 evaluative items and the corresponding 45 intercorrelations, only two of these correlations were not statistically significant. This high level of interrelatedness between the first 10 items echoes the results from the factor analysis – students tend to think unidimensionally about course and instructor evaluation.

The examination of global course evaluation (Item #12), global instructor evaluation (Item #14), student expected grade (Item #11) and student actual grade (Item #15) yield interesting results. As seen in Table 2, course and instructor evaluations are highly correlated ($r = .632$). However, closer examination of course and instructor evaluation finds different moderating variables. For instance, the course evaluation is significantly (but weakly) correlated with expected grade ($r = .207$, $r^2 = .042$) but course evaluation is not significantly correlated with actual grade ($r = .082$). Conversely, the instructor evaluation is significantly (but weakly) correlated with actual grade ($r = .186$, $r^2 = .034$) but instructor evaluation is not significantly correlated with expected grade ($r = .117$). This pattern of results is also reflected in the regression data that follows.

Using a multiple regression approach to predict the global instructor evaluation ("Overall, I would rate this instructor as"), a significant linear relationship was observed amongst the variables with this global item, $F(13, 283) = 23.38$, $p < .001$, $R^2 = 0.518$. There were three statistically significant predictors of the global instructor evaluation question: a) "Overall, I would rate this course as," $b = 0.388$, $t = 7.33$, $p < .001$, partial $r^2 = .16$, b) "I would recommend this instructor to another student," $b = 0.194$, $t = 3.80$, $p < .001$, partial $r^2 = .048$, and c) "Assignments and exams were returned in a timely fashion," $b = 0.086$, $t = 2.12$, $p < .05$, partial $r^2 = .016$.

A multiple regression approach was also used to predict global course evaluation, $F(13, 267) = 25.47$, $p < .001$, $R^2 = .554$. Again, three factors emerged to predict the global course item: a) "I would recommend this instructor to another student" $b = 0.306$, $t = 5.43$, $p < .001$, partial $r^2 = .100$, b) "I expect to receive a grade of" $b = 0.102$, $t = 2.47$, $p < .05$, partial $r^2 = .022$, and c) "Overall I would rate this instructor as" $b = 0.398$, $t = 7.04$, $p < .001$, partial $r^2 = .157$.

DISCUSSION

In examining the pattern of results across this study, it seems that students use different sources of data to arrive at course and instructor evaluations – while using somewhat different sources of data, however, course and instructor evaluation outcomes remain highly related to one another. The best predictors of the global instructor evaluation score are the overall course rating, the recommend to another student question, and the question concerning if assignments and exams are returned in a timely fashion. The relationship between course and instructor is not surprising, and has been noted earlier. For instructors to attempt to impact their instructor evaluations, they may wish to focus on those items generally held of interest to students (to the extent that a student would recommend a course to another student). One example of the specificity of this influence seems to come from assignments and exams being returned in a timely fashion. This seems to be a slightly different perspective from the literature, which has focused on other aspects of influence of instructor evaluations, such as attire (Chowhardy, 1988), animation (Centra, 1993; Williams & Ceci, 1997), and personality (Centra, 1993; Kovacs & Kapel, 1976).

When predicting course evaluations, the three best predictors include the "recommend to another student" question, the expected grade, and the global instructor rating. Given the patterns of results from the regression approach, future research efforts may be fruitful if the factors that most influence a student's recommendation to another student can be identified. In this case, expected grade does predict a course evaluation, as well as the overall instructor rating. These types of results do differ from past studies that have looked at class size (Meredith & Ogasawara, 1981), type of class (DaRosa, et al., 1991; Hoffman, 1983), and difficulty level of the course (Marsh, 1978; Schwab, 1975). Instructors need to be aware of these relationships and design course experiences that meet student needs, while at the same time giving students a fair chance to succeed.

The evaluation form as used in this study indicated good reliability, but in an attempt to establish validity, all 10 specific items loaded on one factor. This outcome again reflects the difficulty that students have in disentangling course evaluations from instructor evaluations. More work in this area with multiple instructors at multiple institutions may be able to demonstrate that the evaluation questions can be used to differentiate course and instructor dimensions.

Whether used for personal improvement or personnel decisions, evaluations need to be carefully used and interpreted. Results from the present study indicate that while course and instructor evaluations seem related, there are subtle differences in those items or factors that influence the outcome of these measures. These results help to update and revisit the issue of student performance and course and instructor evaluations, and continual work in this area needs to be conducted in order to better understand the relationship between course and instructor evaluations, and what factors influence these evaluations.

Address correspondence concerning this article to:

R. Eric Landrum
Department of Psychology
Boise State University
1910 University Drive
Boise, ID 83725-1715

elandru@boisestate.edu

REFERENCES

Centra, J. A. (1993). Reflective faculty evaluation. *Student evaluations of teaching: What research tells us* (pp. 47-79). San Francisco: Jossey-Bass Publishers.

Chowdhary, U. (1988). Instructor's attire as a biasing factor in students' ratings of an instructor. *Clothing and Textiles Research Journal*, 6, 17-22.

DaRosa, D. A., Kolm, P., Follmer, H. C., Pemberton, L. B., Pearce, W. H., & Leapman, S. (1991). Evaluating the effectiveness of the lecture versus independent study. *Evaluation and Program Planning*, 14, 141-146.

Ellis, N. R., & Rickard, H. C. (1977). Evaluating the teaching of introductory psychology. *Teaching of Psychology*, 4, 128-132.

Gigliotti, R. J., & Buchtel, F. S. (1990). Attributional bias and course evaluations. *Journal of Educational Psychology*, 82, 341-351.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52, 1209-1217.

Hoffman, R. A. (1983). Grade inflation and student evaluations of college courses. *Educational and Psychological Research*, 3, 151-160.

Jensen, M. D. (1987). *Ethics, grades, and grade inflation: Student evaluations as a factor in multi-sectioned courses*. St. Louis, MO: Central States Speech Association and the Southern Speech Communication Association. (ERIC Document Reproduction Services No. ED 281 259)

Kennedy, W. R. (1975). Grades expected and grades received—their relationship to students' evaluations of faculty performance. *Journal of Educational Psychology*, 67, 109-115.

Kovacs, R., & Kapel, D. E. (1976). Personality correlates of faculty and course evaluations. *Research in Higher Education*, 5, 335-344.

Liddle, B. J. (1997). Coming out in class: Disclosure of sexual orientation and teaching evaluations. *Teaching of Psychology*, 24, 32-35.

Marsh, H. W. (1978). *Students' evaluations of instructional effectiveness: Relationship to student, course and instructor characteristics*. Toronto, Ontario, Canada: American Educational Research Association. (ERIC Document Reproduction Services No. ED 155 217)

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. The critical issues of validity, bias and utility. *American Psychologist*, 52, 1187-1197.

Meredith, G. M., & Ogasawara, T. H. (1981). Lecture size and students' ratings of instructional effectiveness. *Perceptual and Motor Skills*, 52, 353-354.

Moomaw, W. E. (1977). Practices and problems in evaluating instruction. *New Directions for Higher Education*, 17, 77-91.

Ory, J. C., & Parker, S. A. (1989). Assessment activities at large, research universities. *Research in Higher Education*, 30, 375-385.

Schwab, D. P. (1975). Course and student characteristic correlates of the course evaluation instrument. *Journal of Applied Psychology*, 60, 742-747.

Vasta, R., & Sarmiento, R. F. (1979). Liberal grading improves evaluations but not performance. *Journal of Educational Psychology*, 71, 207-211.

Weigel, R. G., Oetting, E. R., & Tasto, D. L. (1971). Differences in course grades and student ratings of teacher performance. *School and Society*, 99, 60-62.

Williams, W. M., & Ceci, S. J. (1997). How'm I doing? Problems with student ratings of instructors and courses. *Change*, 29(5), 12-23.

Wilson, R. (1998). New research casts doubt on value of student evaluations of professors. *The Chronicle of Higher Education*. Retrieved March 26, 1998, from <http://www.chronicle.com/colloquy/98/evaluation/background.html>

Worthington, A. G., & Wong, P. T. P. (1979). Effects of earned and assigned grades on student evaluations on an instructor. *Journal of Educational Psychology*, 71, 764-775.